Sample Size Requirements in Case-only Designs to Detect Gene-environment Interaction

by Quanhe Yang, Muin J. Khoury, and W. Dana Flanders

Abstract: With advances in molecular genetic technology, more studies will examine gene-environment interaction in disease etiology. If the primary purpose of the study is to estimate the effect of gene-environment interaction in disease etiology, one can do so without employing controls. The case-only design has been promoted as an efficient and valid method for screening for gene-environment interaction. The authors derive a method for estimating sample size requirements, present sample size estimates, and compare minimum sample size requirements to detect gene-environment interaction in case-only studies with case-control studies. Assuming independence between exposure and genotype in the population, the case-only design is more efficient than a case-control design in detecting gene-environment interaction. In addition, the authors illustrate a method to estimate sample size when information on marginal effects (relative risk) of exposure and genotype is available from previous studies.

With advances in molecular genetic technology, genetic markers have been used increasingly in case-control studies to search for gene-environment interaction (1-8). Concerns about selecting appropriate control groups for case-control studies have led to the development of several nontraditional approaches to the study of genetic factors (9). If one's primary interest is to assess possible interaction between genetic and environmental factors in the etiology of a disease, one can use the case-only design which does not require controls. This design has been promoted as an efficient and valid approach for screening for gene-environment interaction under the assumption of independence between exposure and genotype in the population (10-11). To help identify situations in which a case-only design may be preferable to the case-control design, we present a method for estimating the sample size required to detect gene-environment interaction with a case-only study and present sample size estimates for several design scenarios. We also discuss situations in which information on marginal effects of exposure and genotype is available from previous studies.

METHODS

For simplicity, we assume the exposure and susceptibility genotype are dichotomous variables. A key assumption that underlies use of the case-only design to study interaction effects is independence of exposure and genotype in the population (10-13). We also assume that background risk, unrelated to either exposure or genotype, exists and that the disease is rare so that the odds ratio estimates the risk ratio.

Table 1 shows the expected frequency distribution among members of a population according to the presence and absence of exposure and genotype. To calculate sample size, one needs to specify the prevalence of exposure (e), the prevalence of genotype (g), the relative risk for exposure alone (R_e), the relative risk for genotype alone (R_g), the effect of the gene-environment interaction (R_i), the case-control ratio, the type I error (a) and the type II error (ß)(12). As shown by Smith and Day, one may calculate the required sample size for specified values of the odds ratio of interaction (R_i), type I error, and type II error (14) from:

$$\log(R_i) = Z_{\alpha/2}\sqrt{V_N} + Z_{\beta}\sqrt{V_A} \qquad (1)$$

where V_N is the variance of the logarithm of R_i under the null hypothesis, V_A is the corresponding variance under an alternative hypothesis, $Z_{a/2}$ and $Z_{ß}$ are normal deviates which cut off appropriate areas in the tails of the standard normal distribution. For a case-control study, and the notations of Table 1, the results of Smith and Day (14) gave:

$$V_A = \frac{1}{n}\left[\sum_{i=1,2}\left(\frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}\right)\right] \qquad (2)$$

$$V_N = \frac{1}{n}\left[\sum_{i=1,2}\left(\frac{1}{A_i} + \frac{1}{M_i-A_i} + \frac{1}{N_i-A_i} + \frac{1}{T_i-M_i-N_i-A_i}\right)\right] \qquad (3)$$

where A_i is the solution of:

$$R_{MH} = \frac{A_i(T_i - M_i - N_i + A_i)}{(M_i - A_i)(N_i - A_i)} \qquad (4)$$

Because there is no closed formula to calculate expected cell counts under the null hypothesis of no interection, we used Mantel-Haenszel approximation (R_{MH}) to estimate V_N as suggested by Smith and Day (14). The number of cases required with a case-control design is derived by solving equation (1) for n:

$$n = \frac{(Z_{\alpha/2}\sqrt{v_N} + Z_{\beta}\sqrt{v_A})^2}{(\log(R_i))^2} \qquad (5)$$

where $v_N = nV_N$ and $v_A = nV_A$.

We followed a similar approach to derive a formula for the number of cases required with a case-only design to detect gene-environment interaction (R_i). The expected distribution of the cases according to exposure and susceptibility genotype is summarized in Table 2 and the cross product of the data in Table 2 gives R_i. Under the assumption of independence between exposure and genotype in the population, R_i obtained from cases only measures departure from multiplicative joint effect of exposure and genotype (9, 10).

The variance of the logarithm of R_i based on expected values under alternative hypothesis is:

$$V_A = \frac{1}{n}\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right) \qquad (6)$$

We test the null hypothesis $R_i = 1$ with the test statistic $? = \ln(R_N)/\sqrt{V_N}$, where V_N is the null variance V_N. Based on expected values, V_N is calculated from the marginal totals as:

$$V_N = \frac{1}{n}\left(\frac{1}{MN} + \frac{1}{(T-N)M} + \frac{1}{(T-M)N} + \frac{1}{(T-M)(T-N)}\right) \qquad (7)$$

We used the values of V_N and V_A to calculate, using equation (5), the required number of cases in a case-only design to detect interaction given various prevalence of exposure (e) and genotype (g) in the population. For comparison, we also calculated the sample size required in a case-control design.

Estimating sample size when only marginal effects of exposure and genotype are known

In planning a study of gene-environment interaction, one may know the marginal effects of exposure (R'_e) and genotype (R'_g) from previous studies, but not know either the effects of exposure among members of the population who are not susceptible (R_e) or the effects of the susceptibility genotype among the unexposed (R_g). In such a study, one may wish to calculate the number of cases required to detect gene-environment interaction given only e, g, R'_e, R'_g, and R_i. Under the assumption of independence between exposure and genotype in the population, we can do this sample-size calculation because the marginal effects of exposure (R'_e) and genotype (R'_g) are functions of R_e, R_g, R_i, the prevalence of exposure (e), and genotype (g) (13):

$$R'_e = \frac{(1-g)R_e + gR_eR_gR_i}{(1-g) + gR_g} \qquad (8)$$

and

$$R'_g = \frac{(1-e)R_g + eR_eR_gR_i}{(1-e) + eR_e} \qquad (9)$$

By solving equations (8) and (9) for R_e and R_g, we can express R_e and R_g as functions of R'_e, R'_g, and R_i and then use those functions in equation (5) to calculate sample size. Expressions and proof of unique positive solutions for R_e and R_g are in Appendix 1. We also calculated the number of cases required in case-control and case-only designs to detect gene-environment interaction assuming that R'_e, R'_g, and R_i are known.

RESULTS

Sample sizes for various levels of R_e, R_g and R_i of gene-environment interaction

We calculated the required sample size for a range of R_e, R_g, R_i, g and e (Table 3). We present sample sizes for the case-only design and, for comparison, for the case-control design. We present only sample sizes for which g is greater than or equal to e, since the sample size is symmetric with respect to the prevalence of exposure and genotype. Because we used the Mantel-Haenszel approximation to estimate the sample size under the null hypothesis for case-control design (14), the calculated sample sizes for the case-control design are slightly asymmetric, especially with high values of R_i. For example, for $R_i = 10$, $e = 0.3$ and $g = 0.7$ with $g = 0.3$ and $e = 0.7$, the calculated sample size are 93 and 90 respectively. Therefore, for the case-control design we present in Table 3, the average value of the two calculated sample sizes.

As seen in Table 3, the case-only design requires fewer cases than the case-control design to detect interaction. As one would expect, greater interaction (R_i) is associated with increased power to detect interaction, and the required sample size is smallest if the prevalence of exposure and genotype are within the range of 30% to 50%.

Sample size calculation using marginal effects of exposure and genotype

We have also calculated sample sizes based on the marginal effects of exposure and genotype. As seen in Table 4, sample sizes calculated from known or assumed values for marginal effects of exposure, genotype, and gene-environment interaction, also yield fewer required cases for a case-only design than for a case-control design. For $R'_e = 5$ and $R'_g = 2$ in Table 4, changes in R'_e and R'_g have similar effects on sample size requirements as observed in Table 3 for $R_e = 2$ and $R_g = 1$.

Example

Hwang et al.(11) investigated the interaction between maternal cigarette smoking, and transforming growth factor alpha polymorphism on the risk for cleft palate in a population-based sample of infants with birth defects. The distribution of these two risk factors in the study is presented in Table 5. Other studies indicated that about 25 percent of women smoke during pregnancy (15-17). We used values of e = 0.25, g = 0.16 (calculated from Table 5), $R_e = 1$, $R_g = 0.9$, and $R_i = 6.1$ and a case/control ratio of 4 to calculate the required number of cases needed to detect the interaction between TaqI polymorphism and maternal smoking on the risk for cleft palate. We found that 75 cases (375 total subjects) would be needed for a case-control study and that 55 would be needed for a case-only study with power of 0.80.

We next attempted to determine the number of cases required for different values of R_i assuming we know the marginal effects of exposure (R'_e) and genotype (R'_g) from previous studies. From the above example, it can be calculated that $R'_e = 1.5$ and $R'_g = 2$. We assumed that the prevalence of the genotype (g) = 15 percent and that the prevalence of the exposure = 25 percent. Because the prevalence of the genotype (g) is better documented than the value of the exposure (e), we varied the values of R_i and e and calculated the number of cases required for a case-control and for a case-only design. As shown in Figure 1, a case-control study with 100 cases and 200 controls would have low power to detect possible interaction effects with $R_i < 5$, whereas a case-only study with 100 cases would have moderate power if the exposure prevalence were greater than 15 percent.

DISCUSSION

Our results show that the case-only design is more efficient than case-control design to detect gene-environment interaction under the assumption of independence between exposure and genotype in the population. Our findings are consistent with other studies which showed that when the exposure and genotype are independent in the population, the case-only studies produced more precise estimates of the interaction between exposure and genotype than do case-control designs (10-11). The power to detect interaction is associated with increased values of interaction (R_i).

The approach we used to calculate sample size is based on large sample variances. In some extreme situations, for example, a large interaction coupled with a common exposure and genotype, some of the expected cell sizes become very small for the calculated sample size. If any expected cell size is less than five for a given sample size, we suggest recalculating the required sample size for a less extreme situation. For example, one may recalculate sample size assuming a smaller degree of interaction.

The case-only design cannot evaluate an individual's relative risk associated with exposure alone (R_e) or genotype alone(R_g). If the marginal effects of exposure (R'_e) and genotype (R'_g) are available from previous studies, our approach allows one to calculate the sample size required to study interaction.

Although it should typically be the case that exposure and genotype are independently distributed in the population (9), the independence assumption may be violated in some instances. For example, individuals with delayed alcohol metabolism as a result of genetic variation in alcohol aldehyde dehydrogenase may have an increased flushing response after alcohol ingestion (18-19) and thus be more likely to avoid alcohol exposure. In addition, the independent assumption could be contradicted in any population where both the exposure and genotype co-vary with other factors, like ethnicity. Such correlations could also invalidate a case-only design in detecting gene-environment interaction.

The gene-environment interaction (R_i) derived from a case-only design assumes a departure from multiplicative effects. The appropriateness of using such interaction in epidemiologic studies has been discussed elsewhere (20-22). Studies have shown that many biologically plausible modes of gene-environment interaction involve a departure from multiplicative effects (23). If the true underlying model of joint effect is additive, the odds ratio of interaction (R_i) derived from a case-only design may not be an appropriate description of the risk in relation to exposure and genotype (9).

In conducting a case-only study, one should follow the same epidemiologic principles of case selection as one would in conducting a case-control design. A population-based consecutive series of incident cases is ideal. Selection of cases from the general population would be one way help to make the findings of such a study more generalizable.

Researchers are increasingly searching for gene-environment interactions in disease. Examples of such studies include: smoking, TaqI polymorphism, and cleft palate (6-7); lung cancer in relation to debrisoquine metabolic phenotypes (2); glutathione S-transferase class mu, smoking, and sister chromatid exchange (SCE) levels in lung cancer (23); polymorphism at cytochrome p4502E1 with gastric and esophageal cancer due to cigarette smoking and other dietary factors (3); N-acetylation phenotype and bladder cancer (1,5); and cigarette smoking, N-acetylation phenotype, and breast cancer (8). With the rapid advances in molecular technology, one may expect that interest in finding the effects of gene-environment interaction in disease etiology will increase. We believe that, in many instances, the case-only design can be a useful tool with which to rapidly screen for gene-environment interaction.

APPENDIX

Calculation of R_e and R_g as a function of R'_e, R'_g and R_i

The marginal effects of exposure (R'_e) and genotype (R'_g) can be expressed as the function of e, g, R_e, R_g and R_i:

$$R'_e = \frac{(1-g)R_e + gR_eR_gR_i}{(1-g) + gR_g} \qquad (A1)$$

$$R'_g = \frac{(1-e)R_g + eR_eR_gR_i}{(1-e) + eR_e} \qquad (A2)$$

Rearranging equation (A2) for R_g, we have:

$$R_g = \frac{R'_g((1-e) + eR_e}{(1-e) + (1-g)eR_e} \qquad (A3)$$

We now substitute equation (A3) into equation (A1), and solve equation (A1) for R_e, to obtain:

$$\begin{aligned} &R_e^2[(1-g)eR_i+geR_iR'_g] \\ &-R_e[(1-g)eR_iR'_e+geR'_eR'_g-(1-g)(1-e)-(1-e)gR_iR'_g] \\ &-R'_e(1-e)[(1-g)+gR'_g] = 0 \qquad (A4) \end{aligned}$$

R_e is a quadratic function of e, g, R'_e, R'_g, and R_i. If we define:

$a = [(1-g)eR_i+geR_iR'_g]$

$b = -[(1-g)eR_iR'_e+geR'_eR'_g-(1-g)(1-e)-(1-e)gR_iR'_g]$

$c = -\{R'_e(1-e)[(1-g)+gR'_g]\}$

it can be shown that $(b^2-4ac)^{½} > 0$ since $a > 0$ and $c < 0$, hence

$b < (b^2-4ac)^{½}$. Therefore there is one and only one positive solution for R_e. We used positive values of R_e obtained from equation (A4) to calculate R_g using equation (A3). We then used R_e and R_g derived as a function of e, g, R'_e, R'_g and R_i to calculate sample size requirements.

ACKNOWLEDGMENTS

The authors wish to thank Jim Buehler for his helpful suggestions. We also wish to thank Michael Atkinson and Shih-Jen Hwang for their technical assistance; and also two anonymous reviewers for their helpful comments and suggestions on an early draft of this paper.

REFERENCES

1. Cartwright, RA, Glashan RW, Rogers HJ, et al. Role of N-Acetyl transferase phenotypes in bladder xarcinogenesis: A pharmacogenetic-epidemiological approach to bladder cancer. *Lancet* 1982; 2:842-6.

2. Caporaso, N., Hayes RB, Dosemeci M, et al. Lung cancer risk, occupational exposure and the debrisoquine metabolic phenotype. Cancer Res 1989; 49:3675-79.

3. Caporaso N, Landi MT, Vineis P. Relevance of metabolic polymorphism to human carcinogenesis: evaluation of epidemiologic evidence. *Pharmacogenetics* 1991;1:4-19.

4. Shields PG. Inherited factors and environmental exposure in cancer risk. J Occup Med 1993;35:34-41.

5. Hayes RB, Bi W, Rothman N, et al. N-acetylation phenotype and genotype and risk of bladder cancer in benzidine exposed workers. Carcinogenesis 1993;14:675-78.

6. Hwang SJ, Beaty TH, Panny S, et al. Association of transforming growth factor alpha (TGFa) TaqI polymorphism and oral clefts: indication of gene-environment interaction in a population-based sample of infants with birth defects. Am J Epidemiol 1994;141:629-36.

7. Shaw GM, Wasserman CR, Lammer EJ, et al. Orofacial clefts, parental cigarette smorking, and transforming growth factor-alpha gene variants. Am J Hum Genet 1996;58:551-61.

8. Ambrosone CB, Freudenheim JL, Graham S, et al. Cigarette smoking, N-acetyltransferase 2 genetic polymorphisms, and breast cancer risk. JAMA 1996;276:1419-1521.

9. Khoury MJ, Flanders WD. Non-traditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls! Am J Epi 1996;144:207-13.

10. Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only

designs for assessing susceptibility in population-based case-control studies. Stat Med 1994;13:153-62.

11. Begg CB, Zhang ZF. Statistical analysis of molecular epidemiology studies employing case-series. Cancer Epidemiol Biomarkers Prev 1994;3:173-5.

12. Hwang SJ, Beaty TH, Liang KY, et al. Minimum sample size estimation to detect gene-environment interaction in case-control designs. Am J Epidemiol 1994;140:1029-37.

13. Khoury MJ, Beaty TH, Hwang SJ. Detection of genotype-environment interaction in case-control studies of birth defects: how big a sample size? Teratology 1995;51:336-43.

14. Smith PG, Day NE. The design of case-control studies: the influence of confounding and interaction effects. Int J Epidemiol 1984;13:356-65.

15. Windham GC, Swan SH, Fenster L. Parental cigarette smoking and the risk of spontaneous abortion. Am J Epidemiol 1992;135:1394-1403.

16. Fox Sh, Koepsell TD, Daling JR. Birth weight and smoking during pregnancy: effect modification by mother’s age. Am J Epidemiol 1994;139:1008-15.

17. Zhang J, Savitz DA, Schwingl PJ, et al. A case-control study of paternal smoking and birth defects. Int J Epidemiol 1992;21:273-78.

18. Sherman DI, Ward RJ, Yoshida A, et al. Alcohol and aldehyde dehydrogenase gene polymorphism and alcoholism. EXS 1994;71:291-300.

19. Chen CC, Hwu HG, Yeh EK, et al. Aldehyde dehydrogenase deficiency, flush patterns and prevalence of alcoholism: an interethnic comparison. Acta Med Okayama 1991;45:409-16.

20. Greenland S. Basic problems in interaction assessment. Environ Health Perspect 1993;101(suppl 4):59-66.

21. Thompson WD. Statistical analysis of case-control studies. Epidemiol Rev 1994;16:33-50.

22. Rothman KJ. Modern Epidemiology. Boston, MA: Little, Brown and Company, 1986:311-26.

23. Cheng TJ, Christiani DC, Xu X, Wain JC et al. Glutathione S-transferase mu genotype, diet, and smoking as determinants of sister chromatid exchange frequency in lymphocytes. Cancer Epidemiol Biolmarkers Prev 1995; 4(5):535-42.

TABLE 1. Expected distribution of cases for gene-environment interaction analysis in a case-control design*,^

Exposure	Susceptibility genotype	Cases	Controls	Total	Odds Ratio
+	+	a_1	b_1	M_1	$R_eR_gR_i$
-	+	c_1	d_1	T_1-M_1	R_g
(Total)		N_1	T_1-N_1	T_1	
+	-	a_2	b_2	M_2	R_e
-	-	c_2	d_2	T_2-M_2	1
(Total)		N_2	T_2-N_2	T_2	

* Where R_e = disease risk among persons with the exposure without the genotype divided by disease risk among persons with no exposure and no susceptible genotype;
R_g = disease risk among persons with the genotype without the exposure divided by disease risk among persons with no exposure and no susceptible genotype;
R_i = interaction effect of genotype and exposure (ie, factor by which odds ratio for those exposed to e and g is different from the multiplied effects of e and g Individually).

^ For calculation of variance under the null and alternative hypothesis for case-control design, we define,

$a_1 = (geR_eR_gR_i) / 3$ $\quad$ $b_1 = ge$
$c_1 = ((1\text{-}e)gR_g) / 3$ $\quad$ $d_1 = (1\text{-}e)g$
$a_2 = ((1\text{-}g)eR_e) / 3$ $\quad$ $b_2 = (1\text{-}g)e$
$c_2 = ((1\text{-}g)(1\text{-}e)) / 3$ $\quad$ $d_2 = (1\text{-}g)(1\text{-}e)$

e = prevalence of exposure.
g = prevalence of genotype.
$3 = (1\text{-}g)(1\text{-}e) + g(1\text{-}e)R_g + e(1\text{-}g)R_e + geR_gR_eR_i$

TABLE 2. The expected distribution of cases for gene-environment interaction analysis in a case-only design*

Exposure	Susceptibility genotype		
	-	+	
-	a	b	M
+	c	d	T-M
	N	T-N	T

* Where $a = n((1-g)(1-e)) / 3$

$b = n((1-e)gR_g) / 3$

$c = n((1-g)eR_e) / 3$

$d = n(geR_eR_gR_i) / 3$

$N = a + c$

$M = a + b$

$T = a + b + c + d$

$3 = (1-g)(1-e) + g(1-e)R_g + e(1-g)R_e + geR_gR_eR_i$

TABLE 3. Number of cases required for case-control (Nc_c*) and case-only (Nc_o*) studies to detect gene-environment interaction by different scenarios of interaction for 80% power at 5% level of significance, two controls per case

		Re = Rg = 1*										
		Ri = 2*				Ri = 5				Ri = 10		
Prevalence of genotype		Exposure 0.1	0.3	0.5	0.7	Exposure 0.1	0.3	0.5	0.7	Exposure 0.1	0.3	0.5
0.05	Nc_c	4282	1929	1708	2147	526	274	276	390	196	121	138
	Nc_o	2996	1223	997	1167	380	147	123	152	128	52	48
0.10		2293	1041	926	1167	293	154	155	217	115	70	78
		1568	659	552	662	195	84	77	102	66	33	35
0.30			486	442	569		85	89	130		45	54
			308	284	372		51	60	97		29	43
0.50				414	549			103	167			80
				284	400			84	153			79
0.70					757				307			
					600				306			
		Re = 2 and Rg = 1										
0.05	Nc_c	3176	1780	1867	2692	415	269	306	464	167	124	152
	Nc_o	1772	1008	1105	1676	216	123	143	227	73	48	61
0.10		1716	967	1021	1484	236	154	177	271	100	73	91
		939	555	624	964	116	76	95	159	41	34	49
0.30		805	463	506	765	132	90	113	192	65	53	77
		413	281	346	570	59	57	88	168	27	39	74
0.50		736	433	493	779	138	102	147	†	75	74	†
		360	277	368	639	60	77	135	†	34	69	†
0.70		949	571	682	1130	199	162	269	†	116	145	†

0.70	949	571	682	1130	199	162	269	†	116	145	†
	448	385	546	990	88	138	267	†	60	149	†

* Nc_c: number of cases required for case-control study; Nc_o: number of cases required for case-only study.

Re = disease risk among persons with the exposure without the genotype divided by disease risk among persons with no exposure and no susceptible genotype.

Rg = disease risk among persons with the genotype without the exposure divided by disease risk among persons with no exposure and no susceptible genotype.

Ri = interaction effect of genotype and exposure (ie, factor by which odds ratio for those exposed to e and g is different from the multiplied effects of e and g individually).

† Expected cell size here is less than 5, so estimate is unreliable.

TABLE 4. Number of cases required for case-control(Nc-c*) and case-only (Nc-o*) studies to detect gene-environment interaction given exposure (R'e* and R'g*) and relative risk of interaction (Ri) for 80% power at 5% significance level, two controls per case

		R'e = 2 and R'g = 2											
		Ri = 2				Ri = 5				Ri = 10			
Prevalence of		Exposure				Exposure				Exposure			
genotype		0.1	0.3	0.5	0.7	0.1	0.3	0.5	0.7	0.1	0.3	0.5	
0.05	Nc_c	2597	1537	1683	2510	403	301	375	606	185	163	221	378
	Nc_o	1175	774	929	1499	210	159	212	368	104	91	131	241
0.10		1485	853	920	1362	241	164	199	318	113	87	115	194
		694	446	529	846	126	89	116	200	62	49	70	127
0.30			448	459	661		85	91	136		43	48	
			267	301	467		51	60	97		26	32	
0.50				454	637			85	118			43	
				328	494			62	94			32	
0.70					877				151				
					726				131				
		R'e = 5 and R'g =2											
0.05	Nc_c	2318	1793	2384	4139	391	390	571	1041	193	228	358	679
	Nc_o	808	956	1573	3096	157	220	388	794	85	137	255	535
0.10		1296	983	1303	2265	221	208	302	550	108	119	185	351
		470	544	888	1740	90	121	211	429	47	73	135	282
0.30		693	492	647	1128	130	96	133	239	66	51	74	139
		288	309	489	943	56	62	102	203	29	33	58	120
0.50		713	486	634	1108	147	91	117	207	81	46	60	109
		332	335	516	980	68	64	98	189	38	32	51	102
0.70		1017	679	883	1544	227	127	153	260	132	65	73	
		521	504	758	1418	114	95	137	252	68	49	67	

* Nc_c: number of cases required for case-control study; Nc_o: number of cases required for case-only study.
R'e = marginal effect of exposure.

R'e = marginal effect of exposure.
R'g = marginal effect of genotype.
Ri = interaction effect of genotype and exposure (ie, factor by which odds ratio for those exposed to e and g is different from the multiplied effects of e and g individually).
H Expected cell size here is less than 5, so estimate is unreliable.

TABLE 5. Case-control analysis of the interaction between maternal cigarette smoking, the presence of transforming growth factor alpha polymorphism, and the risk for cleft palate*

Smoking	TaqI polymorphism	No. of cases	No. of controls	Odds Ratio	95% CIH
-	-	36	167	1I	
-	+	7	34	1.0	0.3-2.4
+	-	13	69	0.9	0.4-1.8
+	+	13	11	5.5	2.1-14.6

* Data derived from by Hwang et al. (6).
H CI, confidence interval.
I Referent.

Figure 1. Number of cases required for case-only and case-control designs to detect gene-environment interaction by given values of R'e and R'g for 80% power at 5% significance level , two controls per case
(R'e = 1.5 R'g = 2.0 g = 0.15)